Ocean Engineering 124 (2016) 141–156

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

An advanced method for detecting possible near miss ship collisions from AIS data

Weibin Zhang [a], Floris Goerlandt [b], Pentti Kujala [b], Yinhai Wang [a,*]

[a] University of Washington, Department of Civil and Environmental Engineering, Smart Transportation Applications and Research Laboratory, Seattle, WA, 98195 USA

[b] Aalto University, School of Engineering, Department of Applied Mechanics, Marine Technology, Research Group on Maritime Risk and Safety, P.O. Box 12200, Aalto, FI-00076 Finland

ARTICLE INFO

Article history:
Received 31 December 2015
Received in revised form
7 June 2016
Accepted 26 July 2016
Available online 3 August 2016

Keywords:
Near miss
Ship-ship collision
AIS data
Traffic conflict
Maritime safety

ABSTRACT

Maritime accidents have the potential to cause significant financial loss, injury, and damage to the environment. One approach to investigating maritime safety is to focus on near misses, that is, situations which did not lead to an accident but where an accident was narrowly avoided. Based on the principles of the traffic conflict technique, which ranks traffic encounters through a conflict severity hierarchy, this paper proposes a novel model for screening maritime traffic data for near miss ship-ship encounters, particularly for open sea and coastal restricted sea areas. Compared to previous methods, the proposed method has a greater specificity, leaving fewer possible near miss cases to be assessed by navigational experts in a contextualised traffic setting. This is achieved by including the effect of ship size through a ship domain, and by better accounting for the criticality of the encounter direction through the Minimum Distance To Collision concept compared to earlier proposed models. The factors included in the model and their relation are based on expert judgments and using knowledge from previous studies. Model parameters are derived from AIS data points from a reference encounter situation dataset. The developed model has been applied to traffic data from the Northern Baltic Sea. The model is subjected to a number of validity tests, the results of which suggest that the model is adequate for ranking and prioritizing encounters for further assessment in an expert judgment phase to identify near misses. Thus, it establishes a method to enable subsequent research into the validity of near miss information to make statements of maritime safety in relation to collision accidents.

1. Introduction

Marine transport is crucial to economic development around the world, but represents significant financial and safety risk. Though maritime accidents are relatively infrequent, the personal, economic, and environmental costs of accidents can be huge (Heij et al., 2011). Groundings, collisions, and fires are the most frequent maritime accident types globally (Soares and Teixeira, 2001). Specially, ship-ship collisions are one of the most frequently occurring accident types in some locations with high traffic intensities, including the Gulf of Finland and the Singapore Strait (Kujala et al., 2009; Qu et al., 2012). It is therefore of crucial importance for maritime authorities, response authorities, and other stakeholders to have effective tools for analyzing the risks associated with these accidents and for gaining insight in the safety of maritime transportation in different areas.

Many risk analysis models and methodologies have been proposed for analyzing accident risk in general (Ferreira and Couto, 2015) and maritime transportation in particular. For an overview of current research, see Li et al. (2012) and Özbaş (2013) . Additionally, Debnath and Chin (2010) and Goerlandt and Montewka (2015a) describe some recent frameworks for analyzing maritime risk.

Other work does not focus on the accidents per se, but starts from non-accident information to obtain insight in the safety of maritime transportation. Such methods consider the occurrence of certain non-accident events in the traffic system as safety performance indicators. This basic data source for such methods is the data from the Automatic Identification System (AIS), which is a system for information exchange between vessels and between vessels and shore facilities. Focusing here on ship collisions, this approach is followed by Berglund and Huttunen (2009), van Iperen (2012, 2015), Qu et al. (2011), Goerlandt et al. (2012), Wen et al.

Abbreviations: VTS, vessel traffic service; AIS, automatic identification system; TCT, traffic conflict technique; DCPA, distance to closest point of approach; TCPA, time to closest point of approach; BCR, bow crossing range; VCRO, vessel conflict ranking operator

* Corresponding author.
E-mail address: yinhai@uw.edu (Y. Wang).

http://dx.doi.org/10.1016/j.oceaneng.2016.07.059
0029-8018/© 2016 Elsevier Ltd. All rights reserved.

(2015a), Zhang et al. (2015), van Westrenen and Ellerbroek (2015) and Wu et al. (2016).

The work presented in this paper is an extension and improvement on the approach presented by Zhang et al. (2015). These authors propose a model to rank the severity of an encounter between two vessels based on three factors: the distance between the two ships, their relative speed, and the difference between their headings. While this model has been found to be adequate for ranking encounters as a basis for detecting near misses, one challenge with the existing model is that it is not very specific. In other words, the model will result in rather many ship-ship encounters needing to be further manually investigated by re-contextualizing the traffic context (other maritime traffic, environmental conditions, etc.). Such expert judgments can be rather time-consuming, which may impede the practical usefulness of the model.

Considering the above, the specificity of the model proposed by Zhang et al. (2015) is improved in this paper by incorporating the vessel size in the model through accounting for a ship domain. A further improvement is the inclusion of the Minimum Distance to Collision (MDTC) concept, which allows a better distinction of the risk levels at various encounter angles.

The rest of this paper is organized as follows. Section 2 outlines the conceptual basis for analysing the safety performance of vessels in a maritime transportation environment, to give the reader an understanding of the wider context and purpose of the proposed method. Section 3 presents the mathematical method for detecting and ranking potential near miss collisions from AIS data. In Section 4, the method is applied to shipping traffic in the Northern Baltic Sea area, and model evaluation tests are performed to assess the model's performance. Sections 5 and 6 provide some discussion and concluding remarks.

2. Conceptual basis

2.1. AIS data in maritime transportation research

AIS data is increasingly applied as a valuable source of information about ship traffic in maritime traffic engineering and in research addressing the safety of maritime transportation. AIS identifies each vessel equipped with an AIS transmitter, and transmits static and quasi-static data about the vessel (call sign, IMO number, destination, cargo, etc.), as well as frequent updates about the vessel position, speed and course. Such data is stored in shore facilities, from which the maritime traffic in a certain area can be reconstructed and further studied. In recent years, AIS data quality has been improved significantly (Felski et al., 2015; Sang et al., 2015), and further improvements are also possible with proper antenna installation (Last et al., 2015).

AIS data has been adopted in diverse purposes such as studying collision avoidance manoeuvers (Mou et al., 2010), ship trajectory analysis (Xiao et al., 2015), ship speed optimization (Psaraftis and Kontovas, 2014), ship domain analysis (Hansen et al., 2013; Rawson et al., 2014; Wang and Chin, 2015), accident investigation (Wang et al., 2013), and maritime risk models (Mazaheri et al., 2015; Goerlandt and Kujala, 2014; Silveira et al., 2013; Zhang et al., 2013).

It has been identified as one research direction using AIS data to detect possible near miss collisions. Goerlandt et al. (2012) applied the elliptical ship domain proposed by Fujii and Shiobara (1971) as criterion, while van Iperen (2012, 2015) applies a combination of proximity indicators and empirical ship domains to separate near misses from normal encounters. Qu et al. (2011) use fuzzy quaternion ship domains as a basis for counting vessel conflicts in a waterway. Wu et al. (2016) apply different conflict detection mechanisms in a waterway in Texas, including the VCRO method by Zhang et al. (2015).

Zhang et al. (2015) proposes a novel model for near miss ship collisions detection based on AIS data by considering three major factors, distance, relative speed and phase. Unlike the previously mentioned methods, these authors propose an operator which ranks the encounters in terms of their conflict severity, using a continuous ordinal ranking similar to the risk analysis method presented by Debnath and Chin (2010), and Debnath et al. (2011).

2.2. Near miss collisions in the context of traffic conflict research

The detection of near miss collisions has only relatively recently become in focus in maritime transportation research, with early work by Berglund and Huttunen (2009) and a first more elaborate treatment through the nautical traffic conflict technique (NTCT) by Debnath and Chin (2010). This focus on near misses and traffic conflicts has a longer tradition in road safety research, where the traffic conflict technique (TCT) is one of the mainstays, with work by Hydén (1987), Chin and Quek (1997), Svensson (1998) laying early foundations, which have been operationalized and extended also more recently, see Laureshyn et al. (2010) and Zou et al. (2014).

Both the TCT and the NTCT start from the idea that apart from accidents, also other traffic conflicts provide indications and insight into the safety level of the transport system. Such conflicts are situations in which traffic system users encounter one another with the possibility of an accident occurrence, without this accident actually materializing. The underlying reasoning is that conflicts representing a different severity have different safety margin towards an accident occurrence. The hierarchy of different conflict severities includes undisturbed passages, conflicts of various severities, near misses, and actual accidents. This hierarchy has been visually presented as a pyramid (Hydén, 1987) or a diamond (Svensson, 1998). The most significant benefit of the TCT is that vessel conflicts of lower severity levels occur more frequently than collision accidents, so that more can be said about the safety of the transportation system in shorter time periods (Debnath and Chin, 2010). A weakness however is that the exact relation between such vessel conflicts and accident occurrences is not actually very well established yet. Work by Hänninen and Kujala (2014) suggests that there indeed is a relation between near misses (as reported by VTS centres) and accident occurrence, but more evidences and research into this link is desirable. The contribution made in this paper should not be understood to mean that sea areas with more near misses necessarily represent areas of higher collision risk. Instead, the purpose is to define a method to assist experts to judge which encounters qualify as near misses. Those cases then be further linked to relevant accident databases, and other relevant shipping information sources to investigate the relationship between near misses and accidents.

Conflict measures are necessary for the definition of TCT, which are derived from the observable system states at a given time. These states are time-dependent, which corresponds to the intuition that traffic safety dynamically varies as the transportation system evolves over space and time. The conflict measures are used to rank the severity of the vessel encounters on a qualitative ordinal scale, then more serious conflicts can be identified and these can be further connected to the overall traffic safety by ranking the different encounters (Zhang et al., 2015). In almost all traffic conflict methods in maritime traffic, the conflict measures concentrate pairwise encounters between vessels, due to the enormous complexity in accounting for the spatio-temporal relations of multiple interacting traffic users (Laureshyn et al., 2010). An exception is the work by Wen et al. (2015b), who focus on the traffic complexity by explicitly accounting for the density of ships

Please cite this article as: Zhang, W., et al., An advanced method for detecting possible near miss ship collisions from AIS data. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.059

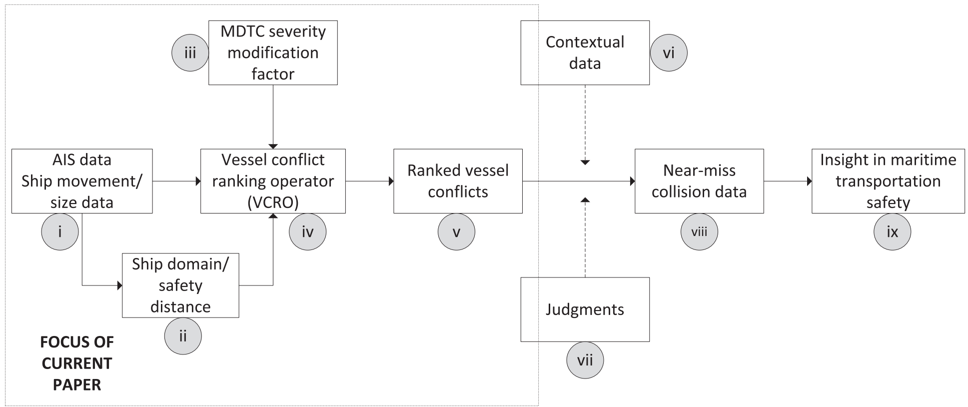


Fig. 1. Framework for research on near miss collisions.

near one another. However, the exact link between complexity and maritime safety is not yet much elaborated upon.

Limiting the modelling ambitions to pairwise encounters, it has been argued that conflict measures from road traffic are not suitable for the maritime transportation system (Debnath and Chin, 2010), which implies a conflict operator suitable for detecting maritime near misses is needed. Furthermore, the limitation to pairwise encounters necessitates a mechanism to look beyond these pairwise encounters, as interactions between multiple vessels can be more relevant from a safety point of view (Zhang et al., 2015; Wen et al., 2015a).

2.3. Conceptual framework for research on near miss collisions

Fig. 1 shows a conceptual framework of how the method developed in this paper fits in the wider context of increasing the understanding of maritime transportation safety, which is very similar to the intended use of the model presented by Zhang et al. (2015). The starting point is AIS data (step i), which provides insights in the spatial and temporal characteristics of traffic in a given sea area. From this data, the safety distance between encountered ships are calculated according to ship size and corresponding ship domain, which is elaborated upon in Section 3.2.1 (step ii). Ship size is also used as a basis of vessels category that reflects ship maneuverability through the MDTC method, see Section 3.2.2 (step iii). Combining these elements which characterize the severity of a navigational conflict, a Vessel Conflict Ranking Operator (VCRO) is defined (step iv.) Subsequently, encounters are detected and ranked (step v) by applying the VCRO in the data (step v).

This is the focus of this paper, especially the formulation of the VCRO to rank encounters for detecting possible near misses encounters. The utility of this method is due to the fact that in large and/or busy sea areas, the number of ship encounters is so high that judging each of these in turn would require too much effort. The VCRO reduces the number of conflicts relevant to consider through a filtering procedure. Compared to earlier work, especially by Zhang et al. (2015) advances are made by making this operator more specific to reduce the number of expert judgments to be made.

The ranked vessel conflicts based on the VCRO (step v) are however not in themselves enough to consider the vessel encounters to be a near miss. As found also by expert interviews reported in Van Iperen (2012), this requires a more extensive contextualization of the encounter, where other factors such as other traffic in the area and the meteorological conditions are considered. This is because navigational situations are experienced as a whole, where navigators interpret the collision risk of an encounter and decide on the necessity of collision avoidance actions based on the proximity of the interacting vessels, their characteristics and the prevailing geospatial and meteorological conditions (Chauvin and Lardjane, 2008; Cockcroft and Lameijer, 2004). Thus, the model is used to suggest the most important encounters to be considered in further judgments, see Fig. 1, rather than to declare the encounter of a certain definite severity. In this regard, the current approach differs from earlier work by e.g. Berglund and Huttunen (2009) and Debnath and Chin (2010).

To achieve this, additional contextual data is added to the highest-ranked vessel encounters (step vi). These are subsequently judged by navigational experts whether or not these qualify as a near miss (step vii). The relevant encounters lead to a database of near-miss collisions (step viii). Steps vi and following are however beyond the scope of this paper.

The database of near-miss collisions can provide insight in the maritime traffic safety (step ix). First, linking the near miss database with accident databases and traffic volumes can provide insight in the actual value of the near miss information in assessing the navigational safety level, e.g. using Bayesian data learning techniques (Hänninen and Kujala, 2014), regression type models (Debnath et al., 2011) or Markov modelling (Faghih-Roohi et al., 2014). This can be used for evaluating the changes to the safety level after design changes in waterways, or for investigating the importance of contextual factors on the occurrence of near misses. These uses are also suggested by Debnath and Chin (2010) and Debnath et al. (2011), who suggest to use traffic conflicts and near misses as diagnostic rather than predictive tools. This issue relates to the validity of the traffic conflict technique, and its theoretical assumption that more severe conflicts do indeed relate to higher accident risk. In road traffic research, the validity of the traffic conflict technique has also been questioned, see Williams (1981) and Hauer and Gårder (1986). A relationship between conflict severity and accidents has indeed been found (Svensson 1998). Research in maritime transportation has however not yet led to conclusive results about the utility of near miss information for analysing collision risk.

3. Defining the vessel conflict ranking operator

The vessel conflict ranking operator (VCRO) VCRO is constructed based on both expert judgment and previous research on maritime safety, and is used to assess the severity of a pairwise vessel encounter. Usually higher ranked encounters represent more severe encounters while lower ranked encounters means safer. During the steps to define the content and structure of the mathematical model of the VCRO, senior nautical officers with more than five years of sea service were interviewed. Therefore, the parameters of the mathematical model are defined based on a combination of expert judgment and parameter estimation using AIS data.

3.1. Model content

As indicated by previous studies of ship collision risk models (Ren et al., 2011; Mou et al., 2010; Bukhari et al., 2013; Goerlandt and Montewka, 2015b), the distance at closest point of approach (DCPA) and time to closest point of approach (TCPA) do not fully reflect the severity level of the vessel encounter. Some studies (Goerlandt and Kujala, 2014; Zhang et al., 2015) have also highlighted the need for more focus on which factors to include in the definition of (the severity of) ship encounters.

First, in certain head-encounters, if the other ship crosses ahead of own ship with a small value of the closest point of approach (CPA) but sufficiently wide bow cross range (BCR), the small CPA would not imply an unsafe encounter. In such cases, the BCR is informative and can be used to make inferences, where the CPA could have led to misclassification of a safe encounter as unsafe, see also Zhang et al. (2015).

Second, the DCPA does not account for the relative orientation of the encountering ships. Encounters in which the other ship comes from abaft of the beam of own ship at an acute angle are known to be more demanding owing to limited manoeuvring space and larger course alteration compared to cases in which the other ship approaches ahead of the beam of own ship and at an oblique or right angle, see also Zhang et al. (2015).

For these reasons, and combined with the expert interviews, the VCRO is constructed using a mathematical model incorporating more generic characteristics of ship-ship encounters by considering following relevant factors:

(1) The safety distance between the two ships. This distance is here not simply to be defined as the distance between the centres of two encountered ships, as e.g. in Zhang et al. (2015), but the distance from the observed ship away from the safety domain boundary of observing ship. This approach is more comparable to work by e.g. Qu et al. (2011), who count overlaps of domains as indicators of ship-ship collision risk.
(2) The rate of change of the distance in the course of the encounter. This is determined by the relative speed of the two ships, and relates to the time available to perform evasive actions.
(3) The relative orientation of the two ships. This is determined by the difference between their headings, and by a factor which accounts for the different intensity of evasive manoeuvring actions. The ship manoeuvrability is a particularly important factor in assessing conflict severity, based on work by e.g. Montewka et al. (2012) and Zhang et al. (2012). As manoeuverability is difficult to assess directly from AIS data, the MDTC model is used as a proxy for this. This factor is based on the MDTC model proposed by Montewka et al. (2010), see Section 3.2.2.
(4) Ship size. As the ship domain is dependent on the ship size, see Section 3.2, the safety distance is indirectly determined by the ship size as well.

These above factors are included as all of these are related to the available time for officers to perform the evasive action and the magnitude of the action required to clear the situation. The time available is directly determined by the distance between the two encountering ships and their relative speed, whereas the magnitude of the action, which relates to manoeuvrability, is assessed based on the difference between the headings of the ships and the MDTC-model. The ship size is related to the ship domain, and hence is incorporated in the distance. These four elements describe the complexity of the encounter between two vessels, and can be used as a basis for ranking the conflict severity in the encounter.

3.2. Model structure

The functional form of the mathematical model of the VCRO can be deduced from the qualitative relations between the individual factors which are found to have a relation to the conflict severity. These are considered in turn, in a slightly different order than presented in Section 3.1, for reasons of clarity and coherence.

3.2.1. Ship domain

A ship domain is defined by Goodwin (1975) as the area around the vessel which the navigator would like to keep free of other vessels for safety reasons. Several attempts have been made and approaches presented for quantifying the size of this domain, see Wang et al. (2009) for an overview.

Domains can be classified by their shape: circular, elliptical and polygonal domains have been proposed. In this paper, the elliptical ship domain proposed by Fujii and Shiobara (1971) is applied. While this is a relatively simple model, it has been empirically confirmed for open waters and coastal restricted waters through data from the AIS system by Hansen et al. (2013). The domain is defined as an ellipse with the major axis along the ship's length and the minor axis perpendicular to the ship's beam, as illustrated in Fig. 2. The half-length of the major axis is taken as $4L$, while the half-length of the minor axis is taken as $1.6L$, with L the ship length. A number of comments are in place regarding the use of this domain, see e.g. Goerlandt et al. (2012):

- The domain is symmetric, which implies that the possible influence of the COLREGs[1] is not taken into account;
- Another consequence of this symmetry is the fact that passing behind the stern is considered as dangerous as passing in front of the bow;
- In the meeting between ships, the largest ship has the largest domain. This means that for the largest vessel, the situation is

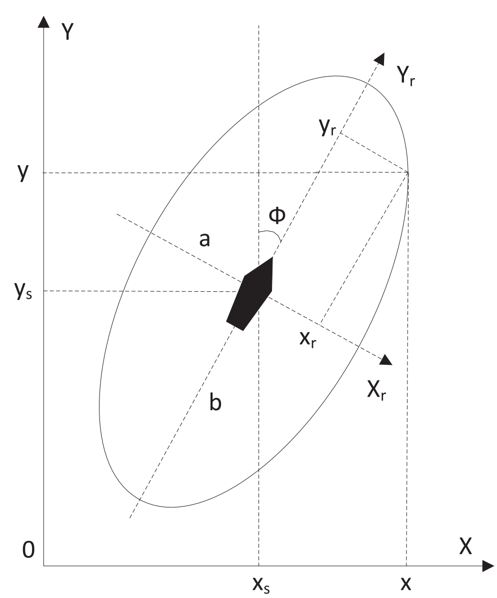


Fig. 2. Ship domain according to Fujii and Shiobara (1971). Figure: Wang et al. (2009). $a=1.6\,L$, $b=4\,L$.

[1] The COLREGs are International Regulations for Preventing Collisions at Sea.

Please cite this article as: Zhang, W., et al., An advanced method for detecting possible near miss ship collisions from AIS data. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.059

classified as dangerous, whereas for the smallest vessel, the situation may still be evaluated as safe.

The above issues justify the view that the ship domain has a certain importance for classifying encounters in terms of their severity, as the violation of domains implies a certain proximity which navigators typically want to avoid. However, the ship domain in itself is not sufficient to assess the severity of conflicts, but should be considered together with other situational characteristics. This will be further considered in Section 3.2.3.

The elliptical domain can be described as (Wang et al., 2009)

$$\begin{cases} f_{ellipse}(x, y)>0, & \text{while } (x, y) \text{ is out of domain} \\ f_{ellipse}(x, y)=0, & \text{while } (x, y) \text{ is on the boundary} \\ f_{ellipse}(x, y)<0, & \text{while } (x, y) \text{ is in the domain} \end{cases} \tag{1}$$

$$f_{ellipse}(x, y)=\begin{cases} \left(\dfrac{x-x_0}{S_t}\right)^2+\left(\dfrac{y-y_0}{R_{t,f}}\right)^2-1, \\ if\, 0^\circ\le\varphi_r\le 90^\circ \text{ or } 270^\circ\le\varphi_r<360^\circ,\ t\in\{b, w\} \\ \left(\dfrac{x-x_0}{S_t}\right)^2+\left(\dfrac{y-y_0}{R_{t,a}}\right)^2-1, \\ if\, 90^\circ<\varphi_r<270^\circ \end{cases} \tag{2}$$

$$\varphi_r=\begin{cases} arccos\dfrac{y_r}{\sqrt{x_r^2+y_r^2}}, & x_r\ge 0 \\ 360^\circ-arccos\dfrac{y_r}{\sqrt{x_r^2+y_r^2}}, & x_r<0 \end{cases} \tag{3}$$

$$\begin{cases} x_r=(x-x_0)\cos\varphi-(y-y_0)\sin\varphi \\ y_r=(x-x_0)\sin\varphi+(y-y_0)\cos\varphi \end{cases} \tag{4}$$

$$\begin{cases} x_0=x_s+d_c\sin(\varphi+90^\circ) \\ y_0=y_s+d_c\cos(\varphi+90^\circ) \end{cases} \tag{5}$$

3.2.2. Minimum Distance to Collision (MDTC) and MDTC-influencing factor c_{MDTC}

The Minimum Distance to Collision (MDTC) is defined as the minimum safety distance for ship collision avoidance in different intersection degrees (Montewka et al., 2010), see in Fig. 3a. The MDTC indicates the minimum distance to collision avoidance and acts as a threshold for collision avoidance. If the MDTC is larger, that means the ships should take the evasive action earlier to collision avoidance, see also Montewka et al. (2012) and Zhang et al. (2012). Larger MDTC values in certain track-crossing indicate more dangerous situation than others track crossing angles, and therefore encounters with a higher MDTC value should have a higher conflict severity and number resulting from the VCRO model.

As seen from Fig. 3a, the MDTC and therefore conflict severity approaches a maximum value when the angle of intersection is in the range of [90,110] degrees. As the MDTC model itself is not available in this research, an approximation and simplification is made here by fitting a curve through the data available in Montewka et al. (2010), see Fig. 3b.

Fig. 3b demonstrates a fitted curve to the data in Fig. 3a based on Fourier series expansion, according to following formalism:

$$c_{MDTC}\sim\sum_{i=1}^{n} m_i\sin(i\bullet z) \tag{6}$$

here, n means the number of Fourier series, and z is the angle of intersection between encountered ships. m_i is coefficients of Fourier series. The X-axis in Fig. 3b represents intersection degree. The number in Y-axis is used as an influencing factor to the VCRO, as will be discussed in Section 3.4. Its value is chosen quite arbitrarily, as it is only a correction factor which should differentiate between different intersection angles, and also because the VCRO model developed in Section 3.3 is an ordinal model aiming at differentiation between different encounter situations. Moreover, in the VCRO model, another coefficient k is derived based on AIS data, see Section 3.4, implying that the actual numeric value of this influencing factor is not so important per se.

The results of the Fourier series curve fitting yield following approximation curve:

$$\begin{aligned} c_{MDTC}\sim & 0.3443\sin z-0.005811\sin(2z)-0.06834\sin(3z) \\ & +0.01177\sin(4z)+0.04933\sin(5z)-0.01347\sin(6z) \\ & -0.002292\sin(7z)+0.01041\sin(8z)+0.01556\sin(9z) \\ & -0.008126\sin(10z)-0.0009892\sin(11z) \\ & +0.007698\sin(12z)+0.001044\sin(13z) \\ & -0.005202\sin(14z)+0.01056\sin(15z) \\ & +0.001526\sin(16z)-0.01129\sin(17z) \end{aligned} \tag{7}$$

The curve defined by Eq. (7) is shown in Fig. 3b, where n equal 17 by sampling accuracy degree.

3.2.3. Distance to safety domain x

Considering that the distance between vessels alone does not describe the complexity of the encounter, it must be used in conjunction with other parameters. However, it is taken that the severity of the encounter reduces with increasing distance between the two ships. Also, the distance is not simply to be defined as the distance between the centres of two encountered ships, but the distance from the observed ship away from the safety domain boundary of observing ship. This is because navigators keep a larger distance between vessels of different sizes, see Section 3.2.1.

In the mathematical model, this relation is accounted for by making VCRO inversely proportional to the distance.

$$VCRO\sim f\left((x-l_\alpha)^{-1}\right) \tag{8}$$

Here x is defined as distance of ships, and l_α is defined as the position of the safety domain considering the course difference α of the ships, see Fig. 4.

Considering Eqs. (1)–(5), there are

$$\alpha=arccos\left(\frac{y_2-y_1}{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}}\right)-\varphi \tag{9}$$

$$l_\alpha=\left(\frac{1+tan^2\alpha}{\frac{1}{S^2}+\frac{tan^2\alpha}{R^2}}\right)^{1/2} \tag{10}$$

$$d=x-l_\alpha \tag{11}$$

Here S = 1.6L, R = 4L.

3.2.4. Relative speed y

The rate of distance change between the two ships is also a relevant factor to near miss collision. As illustrated in Fig. 5, it is calculated according to actual speed of the two ships and their true headings. With experience of senior expert, the higher the relative speed, the less time a crew can address and resolve the

Please cite this article as: Zhang, W., et al., An advanced method for detecting possible near miss ship collisions from AIS data. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.059

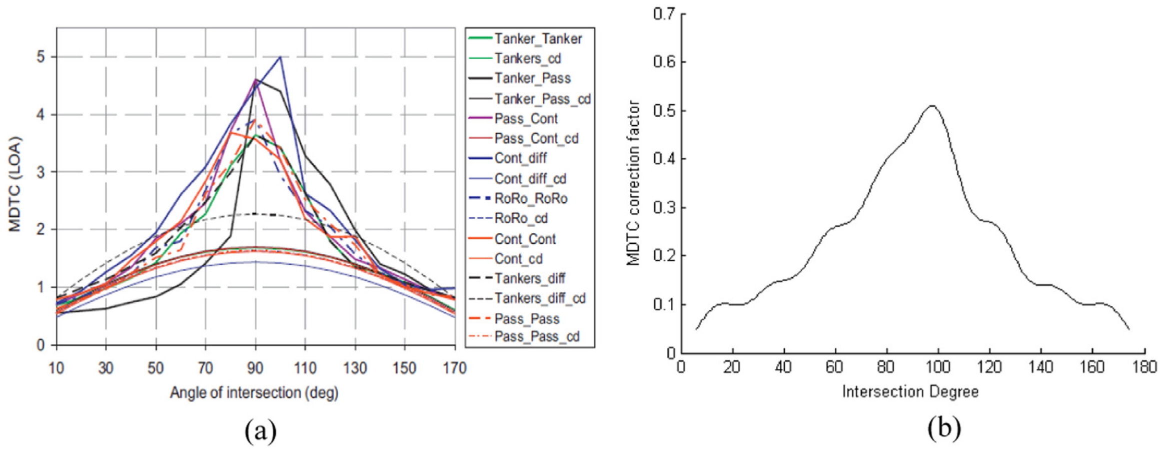


Fig. 3. Values of MDTC obtained for all meeting scenarios, with corresponding values of collision diameters (Montewka et al., 2010).

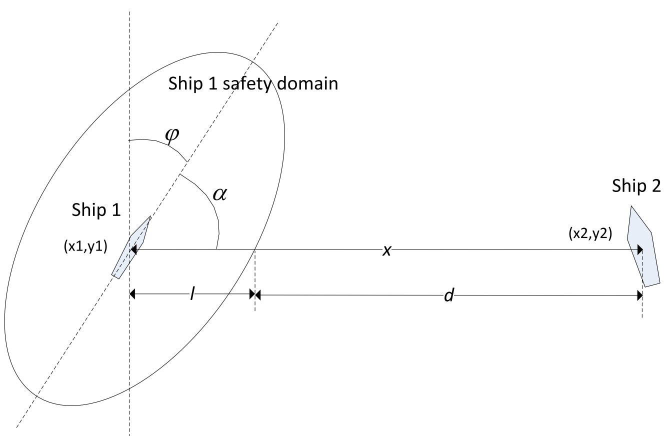


Fig. 4. The distance between in ship pair from the observing ship.

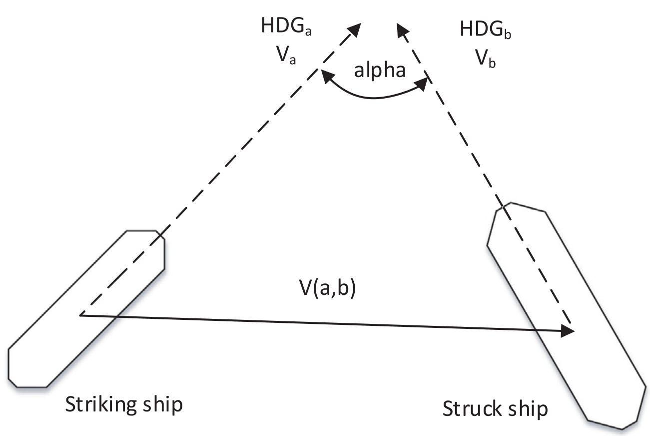


Fig. 5. Definition of the relative speed y.

situation. Therefore, a positive correlation is adopted. The effect of the relative speed is defined below, see also Zhang et al. (2015).

$$VCRO\sim f(y) \tag{12}$$

3.2.5. Phase z

See also Zhang et al. (2015), the phase is defined as the relative orientation of the two vessels. The range of the phase is $[-\pi, \pi]$. A negative value of phase indicates that the encounter is not of interest to navigators, because the ships are away from each other. In contrast, a positive value of phase means that the ships are approaching each other. Since the range of the phase that is in $[-\pi, \pi]$, an odd periodic function $g(z)$ with period 2π is assumed to describe this variable, as follows:

$$VCRO\sim f(g(z)) \tag{13}$$

Fig. 6 illustrates the phase.

In practise, intersection angle of courses between encountered ships has big influence to ship avoidance capacity and manoeuvrability in risk. By the term of phase introduced, the differentiation between encounter scenarios in positive or negative can be distinguished firstly. Secondly, it provides a way to measure a relation between the conflict severity and intersection angle of courses by integrated with c_{MDTC} in Section 3.2.2. c_{MDTC} is used to differentiate the conflict severity, in relation to the interaction angle (i.e. similar to the phase) and also the required manoeuvring intensity to avoid a collision. MDTC used to distinguish the severity of the collision. Therefore, this means that

$$g(z)\sim c_{MDTC} \tag{14}$$

or

$$g(z)=h(c_{MDTC}, z) \tag{15}$$

3.3. Formulation of the mathematical form

Thus, the mathematical functional of VCRO qualitatively has the following form:

$$VCRO\sim f\left((x-l_\alpha)^{-1}, y, h(c_{MDTC}, z)\right) \tag{16}$$

where $g(z)=h(c_{MDTC}, z)$ is an odd periodic function.

A periodic function $\gamma(t)$ with period T can be expanded into a Fourier series as follows:

$$\gamma(t)=\sum_{k=-\infty}^{+\infty} a_k \bullet e^{jk\left(\frac{2\pi}{T}\right)t} \tag{17}$$

In the above, the following notations are adopted:

$$a_k=\frac{1}{T}\int_T \varphi(t)\cdot e^{-jk\left(\frac{2\pi}{T}\right)t} \tag{18}$$

$$e^{j\theta}=\cos\theta+j\bullet\sin\theta \tag{19}$$

Because $g(z)$ is an odd function, it can be described as a sine series:

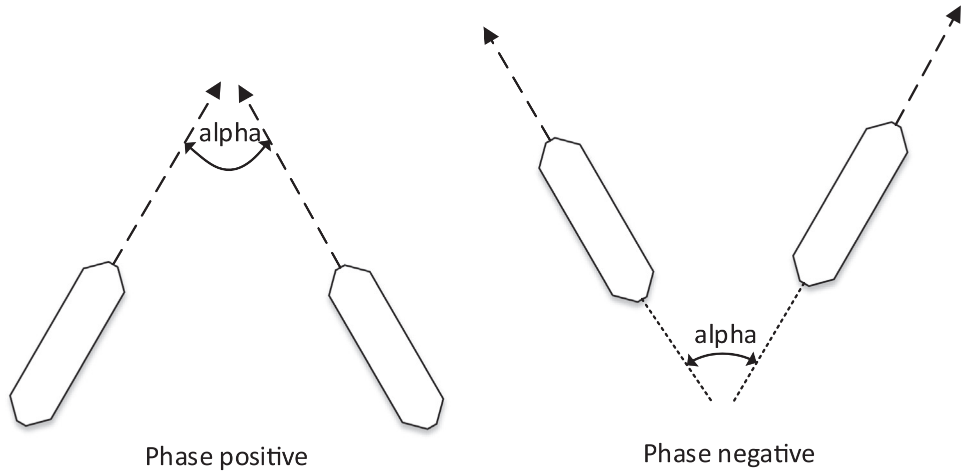


Fig. 6. Definition of the phase z.

$$g_o(x)=\sum_{-\infty}^{+\infty} b_k\bullet \sin\ (kx) \tag{20}$$

In (20), $b_k=a_k\bullet j$, and x represents angle.

Based on the discussion in Section 3.2, especially Eq. (6), VCRO can be expressed as follows:

$$\mathrm{VCRO}(x,\ y,\ z,\ l)=\left(k\bullet(x-l_\alpha)^{-1}\bullet y\right)\sum_{-\infty}^{+\infty} m_i \sin\ (i\bullet z) \tag{21}$$

To reduce the computational complexity, usually a fixed n is adopted in the Fourier series. Previous study (Wang et al., 2014a) indicates, when $n \geq 5$, the series has enough precision to fit a polynomial to an observed periodic pattern. In practise, n is selected according to the sampling density. That is

$$\mathrm{VCRO}(x,\ y,\ z,\ l)=\left(k\bullet(x-l_\alpha)^{-1}\bullet y\right)\sum_{i=1}^{n} m_i \sin\ (i\bullet z) \tag{22}$$

In (22), multiplication is chosen as the mathematical operation for determining the VCRO. This is because the distance and relative speed together determine the conflict severity. As long as the distance is very long and the relative speed is very low, the conflict severity would be very low.

In the special case of zero phase, the ships are moving in the same direction and the effect of the phase on the collision risk would be negligible. The distance and relative speed are thus the dominant factors of the collision risk, and (22), which mathematically leads to a zero value, is replaced by the following equation:

$$\mathrm{VCRO}(x,\ y)=\left(k\bullet(x-l_\alpha)^{-1}\bullet y\right) \tag{23}$$

3.4. Model parameter estimation procedure

A number of typical encounters scenarios are chosen to estimate the parameters k and m_i of Eq. (22) for quantifying the VCRO. Then, the least squares method is adopted to calculate the numerical values of the model parameters. To estimate the parameters, two typical distances are predefined. The 1st distance is 6 NM, which is the usual observation range of a ship-borne radar in on coastal restricted and open sea areas; and the 2nd one is 1 NM away from safety domain. The following conditions of the VCRO are established for the two distances, also see Zhang et al. (2015):

- The absolute value of VCRO for a distance of 6 NM should be low, and is fixed at 5:

$$E\left(\mathrm{VCRO}(x,\ y,\ z,\ l)\right)|_{x=6\mathrm{NM}}=5 \tag{24}$$

- The absolute value of VCRO for a distance of 1 NM from safety domain should be significantly higher, and is fixed at 100:

$$E\left(\mathrm{VCRO}(x,\ y,\ z,\ l)\right)|_{x=1\mathrm{NM}+\text{safety domain}}=100 \tag{25}$$

where E denotes mathematical expectation. The meaning of values (5 and 100) is to recognise encounters of different conflict severity through an ordinal ranking, therefore the numerical value have no importance or significance in themselves.

AIS data sampled from over 2000 different encounters in May to July of 2011 are conducted into the two conditions above to calculate coefficients. The method of least squares is applied to estimate the parameters, as follows:

$$\begin{aligned} f_0 &= \min\left(\sum \Delta\mathrm{VCRO_i}\right) \\ &= \min\left(\left(\sum_i^M k(x_i-l_\alpha)^{-1}y_i \sum_{j=1}^n m_j\ \sin(j\cdot z_i)\right)^2\right) \end{aligned} \tag{26}$$

The above Eq. (26) is used to estimate the parameters k, m_j, $j = 1,2\ldots,n$, from sampled encounters M as detected from AIS data. n is the number of Fourier series. The results are shown in Table 1,

Table 1
Coefficient calculation results.

k	m1	m2	m3	m4	m5	m6	m7	m8
23.22	0.3443	−0.005811	−0.06834	0.01177	0.04933	−0.01347	−0.002292	0.01041
m9	m10	m11	m12	m13	m14	m15	m16	m17
0.01556	−0.008126	−0.0009892	0.007698	0.001044	−0.005202	0.01056	0.001526	−0.01129

which leads to a formulation of the resulting model as follows:

$$
\begin{aligned}
VCRO=&23.\,22(x-l_\alpha)^{-1}\bullet y\bullet[0.\,3443\sin z-0.\,005811\sin(2z)\\
&-0.\,06834\sin(3z)+0.\,01177\sin(4z)+0.\,04933\sin(5z)\\
&-0.\,01347\sin(6z)-0.\,002292\sin(7z)\\
&+0.\,01041\sin(8z)\\
&+0.\,01556\sin(9z)-0.\,008126\sin(10z)\\
&-0.\,0009892\sin(11z)\\
&+0.\,007698\sin(12z)+0.\,001044\sin(13z)\\
&-0.\,005202\sin(14z)\\
&+0.\,01056\sin(15z)+0.\,001526\sin(16z)\\
&-0.\,01129\sin(17z)]
\end{aligned}
\tag{27}
$$

where x is distance, l_α is calculated by Eq. (10), y is relative speed., z is phase. Usually $n \geq 5$ (Wang et al., 2014a) to make sure enough accuracy of result, and n equals to 17 in this study for sampling accuracy degree limitation.

4. Results and model evaluation

In this section, some results of the model application are shown. For this, a distinction is made between various ship size classes and ship-ship encounter classes, to provide an insight into which ship size classes are comparatively more involved in the different conflict severity levels. Moreover, the model is evaluated by subjecting it to two tests, namely the discriminant and concurrent validity tests. Selected encounter scenarios are used for this purpose, and a spatial analysis is made of the locations where encounters of different severity levels are found to occur. The AIS data applied in this study has been obtained from HELCOM (2012).

4.1. Ship size and encounter categories

As indicated in Section 3.1, the ship size category plays an important role during encounters, mainly because of the larger area needed for larger vessels to manoeuver. In presenting the results, a distinction is therefore made between different ship size classes and ship-ship encounter categories. Here, the vessel size dimensions of the ships navigating in the Northern Baltic Sea are extracted and clustered. Fig. 7 demonstrates the results of this analysis.

In Fig. 7, the total number of ships is 2365, from May to July in 2011, which does not include the vessels operated in ports such as tugs and pilot vessels, as well as passenger vessel with size less than 35 m. Clustering method (Wang et al., 2014b) is used to category the ship according to their size, including both ship length and breadth in the studied area. About the clustering method, please reference the description in Section 4.3. By the clustering result, the ships can be divided as three categories as large, medium, and small. The result is shown in Table 2.

Using the ship size categories of Table 2, the encounter types can be divided into 6 categories as shown in Table 3. These are used in the subsequent sections to show the results.

4.2. Model evaluation: discriminant validity

In this section, a number of validity tests are performed to validate the VCRO model. These tests are used to evaluate whether distinguishes situations representing a different conflict severity level can be recognized the model appropriately (Trochim and Donnely, 2008).

4.2.1. Encounter scenarios

For this test, six encounter scenarios sampled in AIS data of May to July 2011 in the Northern Baltic Sea area are chosen from the encounters database, and their resulting VCRO values calculated. These scenarios are shown in Figs. 8–10. Each encounter is described on six sub-pictures in order of Safety Distance, Relative Speed, Phase, the trajectories of the vessels, VCRO in the 1st ship, and VCRO in the 2nd ship. The trajectories of the vessels engaged in the encounter, where 'x' marks the starting locations of each vessel, 'o' indicates their final locations, and red '*' indicates the location in the minimum distance. The curve composed with 'x' signifies the trajectory of the 1st ship, and the curve composed with '+' shows the trajectory of the 2nd ship. The lower panel shows the evolution of the variables distance x, relative speed y and phase z over the duration of the encounter. It also shows the corresponding value of the VCRO. The time axis is a numerical representation of the timestamp of the AIS message. A number of senior nautical officers are interviewed whether the resulting rankings are in line with their judgment as to which of the encounters are more severe.

In Fig. 8 scenario A, two vessels, IMO9274800 (The IMO number is a unique 7-digit number assigned to sea-going merchant ships under the International Convention for the Safety of Life at Sea (SOLAS)) is a tanker with 249 m length and IMO9345398 is a cargo with 205 m length, are interacting in a crossing encounter, where both vessels maintain course and speed at with a relatively large minimum passing distance of 1.3 NM and low relative speed. The

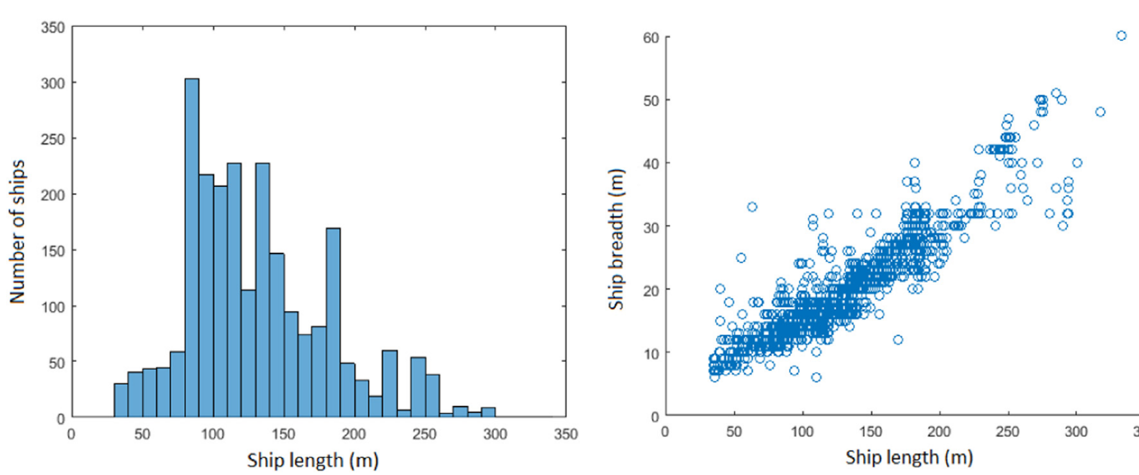


Fig. 7. Ship size analyses in the Northern Baltic Sea.

Table 2
Ship size clustering results.

Category	Ship length	Ship breadth	Average	Number	Ratio
Small	[35,124]	[6,33]	91.82/14.22	1226	51.84%
Medium	[124,196]	[12,40]	155.66/22.76	892	37.72%
Large	[197,333]	[24,60]	237.18/36.56	247	10.44%

Table 3
Ship encounter statistic in the Northern Baltic Sea during May to July in 2011.

Encounter type	Number	Ratio
Small–Small	17,354	25.41%
Small–Medium	23,463	34.36%
Small–Large	8330	12.20%
Medium–Medium	9905	14.50%
Medium–Large	7656	11.21%
Large –Large	1587	2.32%
Total	68,295	100%

VCRO value is 17, which is a low value. In Scenario B, it is seen that two vessels (IMO9322554, IMO9256327) encounter nearly crossing courses at beginning with a large distance and with a high relative speed. Then one vessel changes the course. Both vessels are cargo vessels, and the lengths are 141 m and 161 m separately. The maximum VCRO during the counter is 50, which still is a low value. These low values are in line with the judgments of the nautical experts, as these encounters are uneventful and represent an undisturbed passage or an evasive manoeuver with low conflict severity.

Fig. 9 Scenario A shows a crossing encounter (IMO9212486, IMO9341108) where the minimum safety distance is about 0.27 NM. This encounter results in a VCRO of 80, which is a relatively high value. According to the nautical experts, these encounters are still within the bounds of normal operational practice, but given the fact that there is less room for error than in the earlier shown scenarios, the higher values of the VCRO are warranted. Scenario B shows a more complex encounter, where it is seen that two vessels (IMO8027638, IMO8408442) encounter one another on nearly reciprocal courses at the beginning with a large distance and with a high relative speed. During the fast approach between these two vessels with minimum passing safety distance of 0.51 NM, both change course. Both vessels are cargo vessels, and the lengths are 96 m and 95 m separately. The maximum VCRO during the counter is 105, which is a high value.

In Fig. 10 scenario A, two vessels, IMO9436240 is a cargo vessel with 120 m length and IMO8803769 is a passenger vessel with 203 m length, are interacting in a crossing encounter, where one vessel maintains course and speed while the other one performs a normal evasive manoeuver at with a relatively small minimum passing distance of 0.06 NM. The VCRO value is 160, which is a higher value than in the previous scenarios. In scenario B, two vessels (IMO9188506, IMO9113018) are engaged in a crossing encounter, then one vessel change direction severely. Both vessels are cargo vessels, and the lengths of both are 118 m. The minimum passing distance is 0.15 NM. The VCRO in this encounter is very high with a value of 690.

As can be seen from Figs. 8 to 10, the encounters with a higher value of VCRO correspond to encounters with less room for error than encounters with a lower VCRO-value. According to the traffic conflict technique, such encounters represent a higher conflict severity and are more likely to be assessed by experts as a near miss, if the wider context of the encounter is accounted for (other traffic in the area and environmental conditions), as described in Section 2.3. Thus, these discriminant validity tests the developed model serves its purpose.

4.2.2. Comparison of discriminant validity between proposed and previous model

One of the main reasons for making an updated model compared to a previously presented model, is to make a model which better discriminates encounters, in order to reduce the number of cases to be further subjected to expert judgment, see Section 2.3. Comparing with the previous model, the new model takes vessel size into consideration through the ship domain, and accounts for the intensity of evasive manoeuvring in different encounter angles through the MDTC model. The old previous model has following form (Zhang et al., 2015):

$$\mathrm{VCRO}(x, y, z) = \Big(\big(kx^{-1}y\big)(m\bullet \sin\ (z) + n\bullet \sin\ (2z)\Big) \tag{28}$$

From Eq. (28), we see that the vessel size and the MDTC influencing factor are not included in the model. The differences between the VCRO as determined in the previous and currently proposed VCRO are demonstrated in Fig. 11. Each encounter is described on seven sub-pictures showing subsequently Safety Distance *x*, Relative Speed *y*, Phase *z*, the trajectories of the vessels, VCRO for the 1st vessel, VCRO for the 2nd vessel, and VCRO according the previous model.

In Fig. 11A, two vessels, IMO8853946 is a bunkering tanker of Estonia with 57 m length and IMO9354284 is a passenger vessel with 212 m length, are interacting in a crossing encounter, where one vessel maintains course and speed while the other one performs a course change after the encounter. The old model indicates the VCRO is 79 whereas the new model indicates a VCRO value of 72 for IMO8853946, and 1325 for IMO9354284. The minimum passing distance is −0.04 NM that means the minimum distance is less than the safety domain.

In Fig. 11B, two vessels, IMO9381380 with 120 m length and IMO8912027 with 87 m length both are cargo vessels, interacting in a crossing encounter. The relative speed decreases fast during the encounter, and the minimum safety distance is 0.05 NM. The old model indicates the VCRO is 327 whereas the new model indicates a VCRO value of 1389 for IMO9381380 and 566 for IMO8912027.

From Fig. 11, we can conclude the following:

(1) The new VCRO proposed here is more reasonable than previous VCRO because the size of ship in taken into consideration, therefore it can indicate cases where the distance between ships is shorter than ship domain.
(2) The VCRO of two encountering vessels may be different because the different ship sizes will affect the respective ship domain sizes and hence the safety distances. In judging the encounter severity, it is advisable to retain only the maximum value of both VCRO values.
(3) From the examples, we can see that the new model of VCRO leads to higher numerical values than the previous model, which is beneficial from a viewpoint that the model should clearly discriminate encounters with different severities. The higher values are beneficial especially if the results of the VCRO analysis are clustered into different groups of conflict severity, e.g. to show the encounters with different encounter severity on a chart.

4.3. Model evaluation: concurrent validity

As another evaluation of the VCRO model to detect potential near misses by ranking vessel encounters is performed using a concurrent validity test. To execute this test, calculation of the VCRO value for detected encounters in AIS data of May to July 2011

Please cite this article as: Zhang, W., et al., An advanced method for detecting possible near miss ship collisions from AIS data. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.059

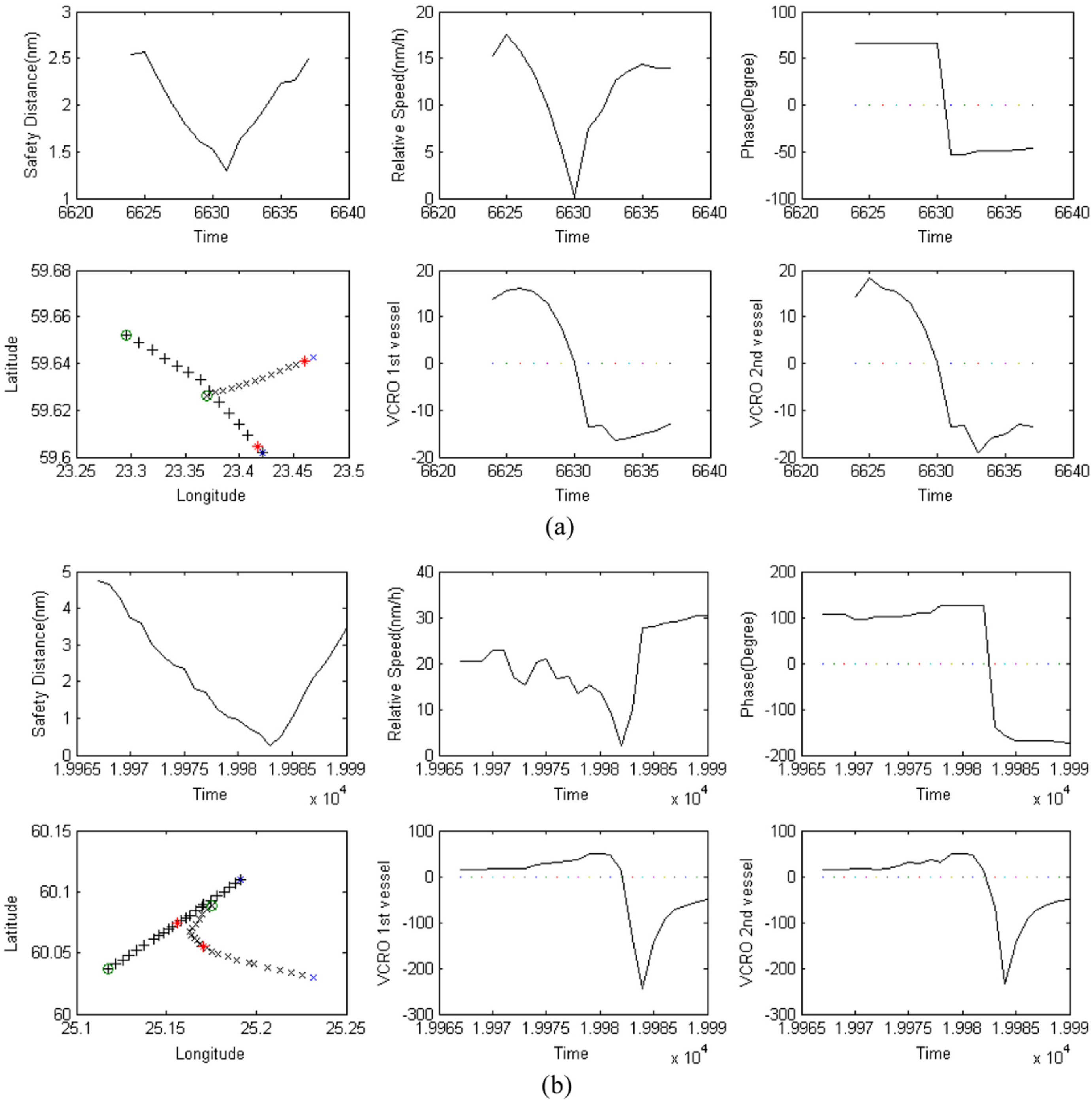


Fig. 8. Encounter scenarios with low conflict severity. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

in the Northern Baltic Sea area are performed. VCRO value grouped in clusters represent different levels of conflict severity, and were plotted on a map. Subsequently, this spatial image demonstrates different collision severity recognition and compared with results of a maritime risk analysis for this area.

The k-means clustering method is performed on the VCRO-values. This is a method of aiming to partition *m* observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster (Hartigan and Wong, 1979). The algorithm is described as follows.

Given a set of observations (x_1, x_2, …, x_m), where each observation is a D-dimensional real vector, k-means clustering aims to partition the *m* observations into *k* sets ($k \le m$) $S=\{S_1, S_2, \ldots, S_k\}$ so as to minimize the within-cluster sum of squares (Zhang et al., 2015):

$$\arg\min \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \tag{29}$$

where μ_i is the mean of points in S_i.

The spatial image of clustered VCRO results is shown in Fig. 12, whereas a summary of the clustering results is shown in Table 4. Cluster 1 means all encounters with negative VCRO, whereas all positive VCRO are clustered into 4 clusters from cluster 2 to 4 using Eq. (29).

The areas where the highest ranked encounters (cluster 4 and

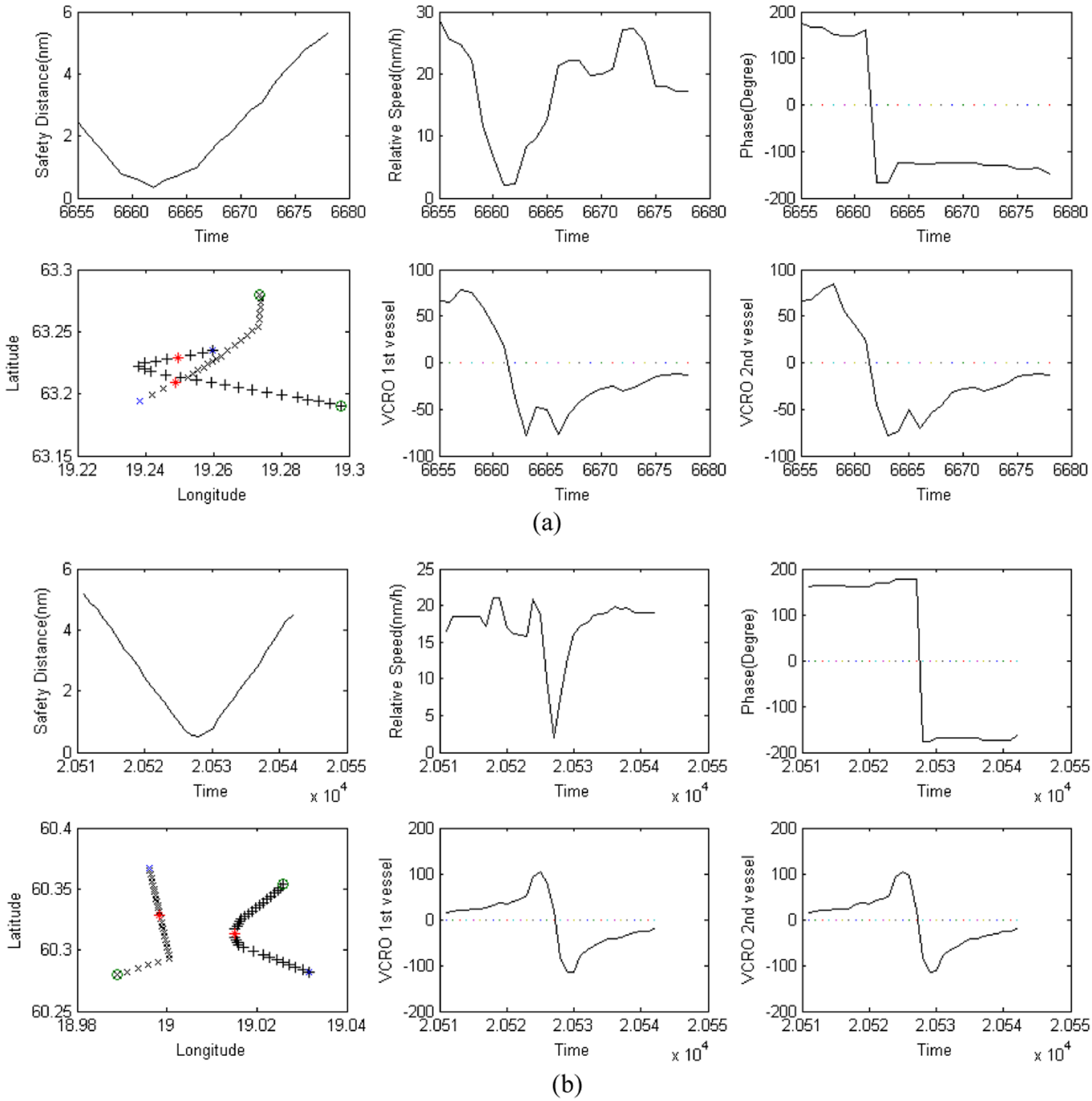


Fig. 9. Encounter scenarios with medium conflict severity. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

5) occur are located in the waterway crossings between Helsinki and Tallinn, the sea area off Stockholm, the waiting area off Kronstadt in the eastern Gulf of Finland and the Northern Quark strait. In the other sea areas, there either are almost no encounters detected, or these are ranked with low or medium conflict severity (cluster 1–3). This is consistent rather well with a previous risk analysis study performed in the Baltic Sea (COWI, 2012), which identifies the same areas listed above as cluster 4 and 5 areas as areas with a high ship-ship collision frequency. Assuming that a higher number of possible near misses does indeed correlate with a higher accident frequency, this concurrent validity test provides some warrant that the near miss detection method is reasonable. However, is should be acknowledged that this relationship between near misses and accidents is not empirically established as outlined in Section 2.3. In addition, the reliability of ship collision risk models has been found to be rather low (Goerlandt and Kujala, 2014), with important uncertainties underlying the model construction. This implies that the areas of high collision risk in Fig. 12(b) may not be correctly identified. Hence, the correspondence seen in Fig. 12 between areas of possible near misses and calculated collision frequencies should not be given too much weight. It merely suggests that a relationship is plausible, but further research is needed to more conclusively investigate this.

Moreover, a number of high VCRO values (cluster 4 and 5) near harbour is mostly related to the close distances at which ships pass each other on opposite courses. In such scenarios, it should be

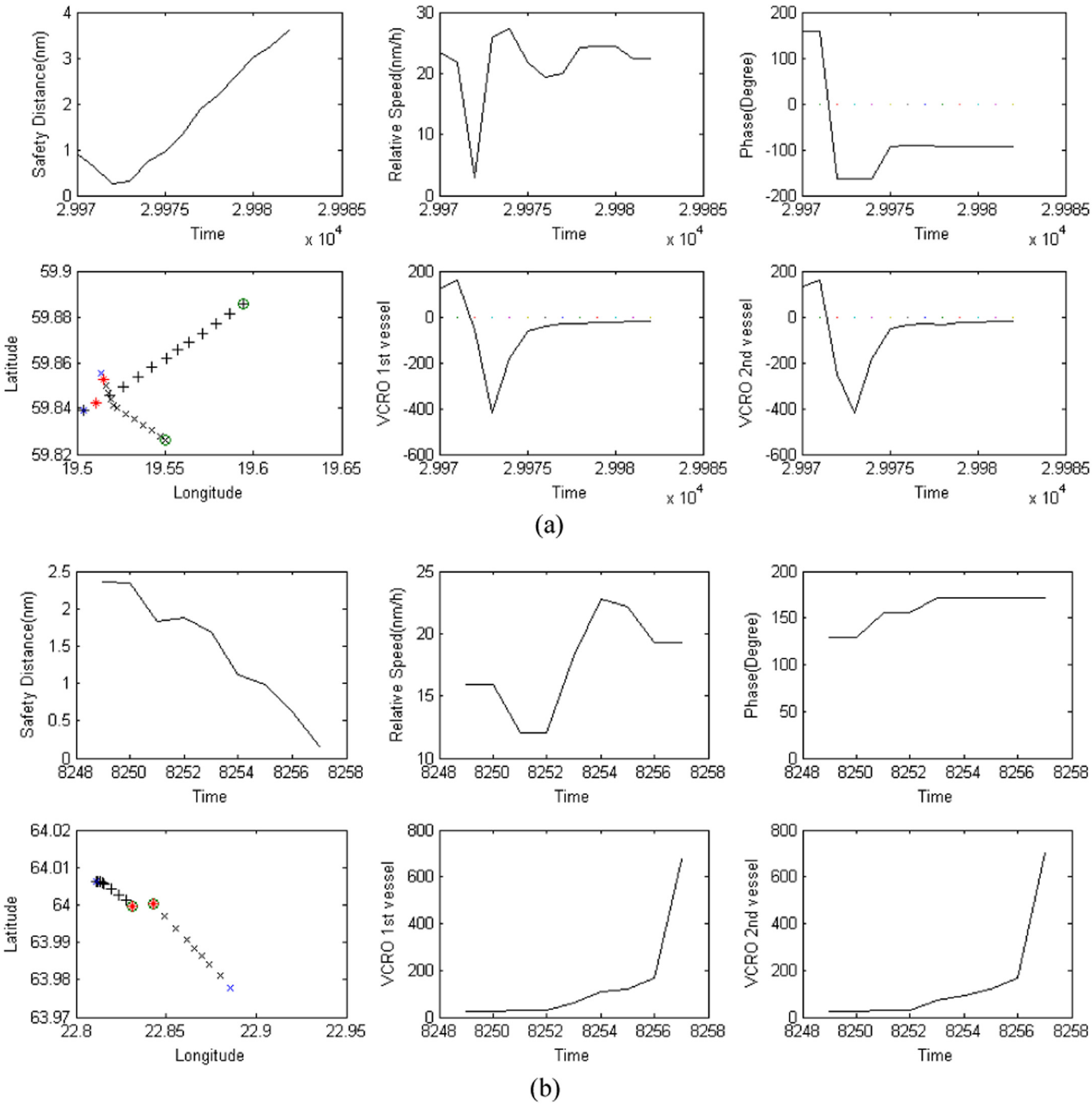


Fig. 10. Encounter scenarios with medium-high conflict severity. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

noted here that the model is intended to be used in open sea and coastal restricted waters, due to the chosen ship domain model which has been empirically validated for these areas, see Hansen et al. (2013). If different ship domains are established for waterways and port approaches, the proposed model can be further modified to account for these.

The model utility can be appreciated from Table 4, where it is found that cluster 4 encounters account for only 1.52% and class-5 encounters account for only 0.09% of all 68,295 encounters, i.e. a significant reduction of cases to be further inspected to find near misses. However, in actual applications of the model, it may be advisable to limit the study area to coastal restricted and open sea areas excluding harbour approaches.

Most of the cluster-3 encounters occurred in waterways to harbours. The fact is reasonable because the navigable space close to harbours is usually limited. The spatial distribution of the cluster-1and 2 encounters took place almost random in the Gulf of Finland. However, their occurrence is correlated to major shipping lanes.

As indicated in Table 4, cluster-5 encounters constitute 0.09% of all encounters. Cluster-4 encounters constitute 1.52%, cluster-3 encounters constitute 12.59%, and cluster-2 and cluster-1 encounters constitute 85.8% together. This distribution of encounters of different severity levels is in agreement with the pyramid

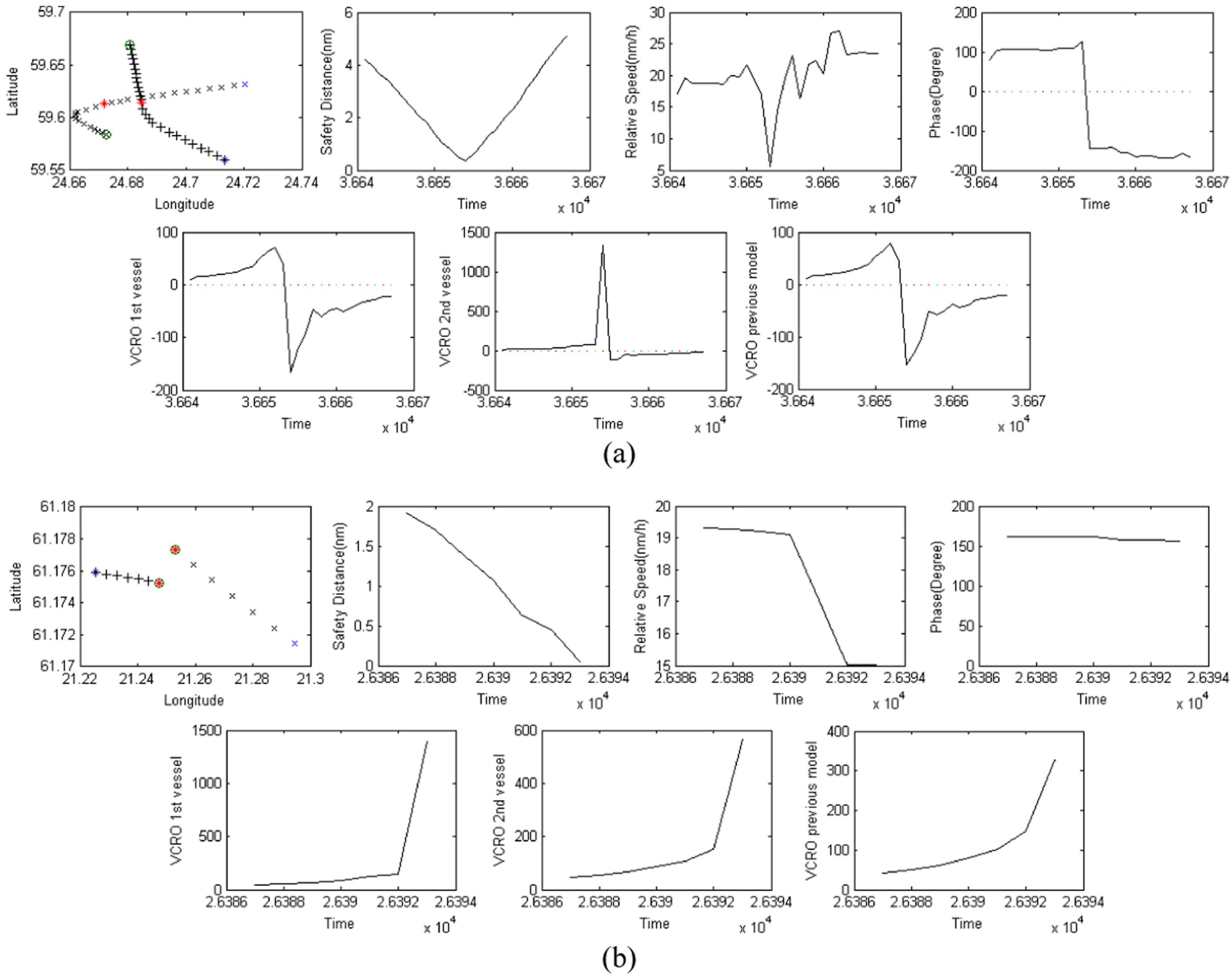


Fig. 11. Difference between models by considering safety domain.

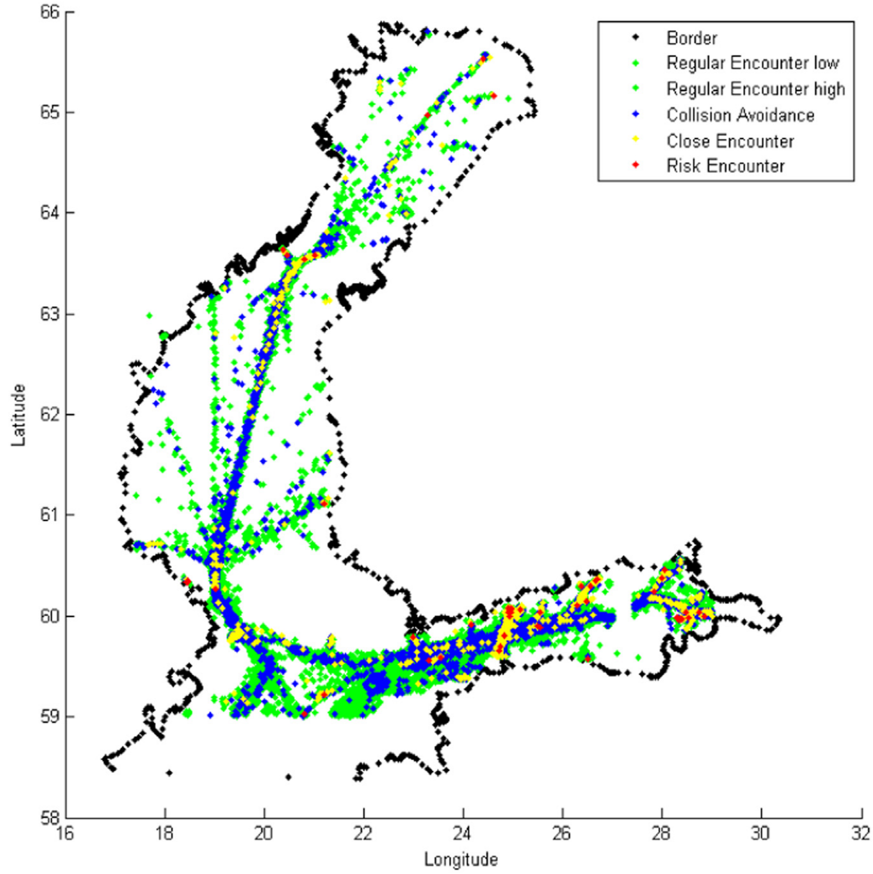
hierarchy found also by previous studies (Hydén,1987; Chin and Quek, 1997; Laureshyn et al., 2010; Debnath and Chin, 2010).

Table 5 shows the performance comparison between VCRO model in this study and previous VCRO model (Zhang et al., 2015). The new VCRO model recognizes 63 encounter scenarios with highest conflict severity from 68,295 cases in total that is 0.09%, whereas previous VCRO model recognizes 303 encounter scenarios from 25,421 cases in total that is 1.2%. By the comparison, the new VCRO improves performance to recognize the potential highest conflict severity scenarios.

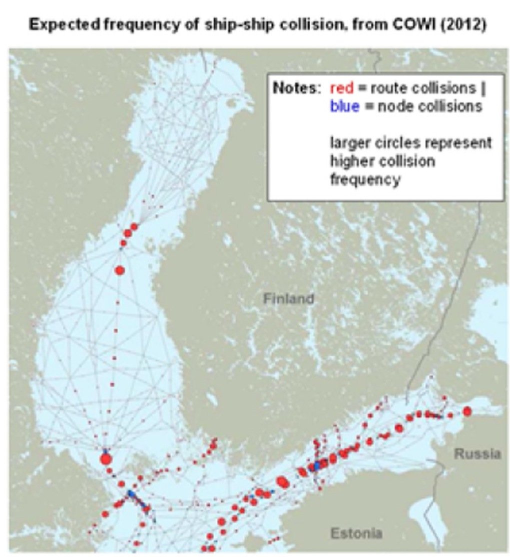
5. Discussion

The model proposed in this paper studied how to detect a near miss ship-ship encounter on a coastal restricted and open sea area. Comparing to a related previous study (Zhang et al., 2015) which combines three variables (the distance between encountering ships, the relative speed of the ships and the difference between the ship headings), ship size and Minimum Distance to Collision (MDTC), which is related to the urgency of evasive manoeuvring are added to the enhanced model. Furthermore, the analysis utilises information derived from combination of actual encounters and expert knowledge to define the model parameters. In addition, the model can be used to rank ship encounter conflict without experts involved. However, the judgement whether the detected encounters qualify as near misses still needs be based on expert knowledge, taking into account additional contextual factors (environmental conditions, other nearby traffic, etc.) after considering the conflict ranked by the proposed model. Hence, the VCRO model does not automatically lead to a set of near misses. Moreover, as stated in Section 2.3, the identification of near misses is only a necessary first step to establish whether or not a relation between near misses and accidents exists.

The first issue is the difference between sizes of ship encountered. In fact, a main improvement in this study is that the size (which relates to manoeuvrability), of the encountering ships is taken into consideration. Ship size affects the distance at which passing is considered safe, see e.g. Wang et al. (2009) and Hansen et al. (2013). However, the safe passing distance for ships of very good manoeuvrability like RoPax is most probably shorter than that for large and bulky ships like tankers, as discussed e.g. in van Iperen (2012) and Goerlandt and Montewka (2015b). This was not be incorporated in the current model, but it could further refine the specificity of the ranking. One way to incorporate this effect would be to make the safety domain dependent also on ship type. However, there currently are no empirical studies available where the ship domain is studied for different ship types, although



(a)



(b)

Fig. 12. Concurrent validity test: VCRO method (left) compared with risk analysis (right).

improved AIS data warehousing and data mining techniques make this possible (Tsou, 2014).

The second issue is the relationship among the parameters. The relationship between the VCRO and the safety distance, phase and MDTC-related effect is relatively clear and has been discussed. In earlier chapters however, the relationship between the VCRO and the relative speed is equally well established. In the model, a linear dependency is assumed, whereas there is no very strong basis for this. Further studies could therefore be performed to investigate alternative hypotheses and check the sensitivity of the rankings to this variable. Considering the evaluation results in Section 4, the linear relationship seems to be rather reasonable, as the model of Eq. (27) based on this relationship, does give rather good results.

The third issue is that the effects of hydro-meteorological conditions such as wave height and visibility were also not taken into consideration in the model. Related models addressing ship-

Table 4
Overview of the number of encounters in each cluster.

Encounter cluster		**Cluster 1** negative VCRO-values no conflict severity	**Cluster2** low VCRO-values low conflict severity	**Cluster 3** medium VCRO-values medium conflict severity	**Cluster 4** high VCRO-values high conflict severity	**Cluster 5** highest VCRO-values highest conflict severity	**Total**
Small-Small	Number	3323	12,727	1198	77	29	17,354
	Ratio	19.15%	73.34%	6.90%	0.44%	0.17%	
Small-Medium	Number	4524	14,704	3707	522	6	23,463
	Ratio	19.28%	62.67%	15.80%	2.22%	0.03%	
Small-Large	Number	1693	5864	733	32	8	8330
	Ratio	20.32%	70.40%	8.80%	0.38%	0.10%	
Medium-Medium	Number	1751	6092	1726	324	12	9905
	Ratio	17.68%	61.50%	17.43%	3.27%	0.12%	
Medium-Large	Number	1289	5269	1024	67	7	7656
	Ratio	16.84%	68.82%	13.38%	0.88%	0.09%	
Large-Large	Number	287	1072	208	19	1	1587
	Ratio	18.08%	67.55%	13.11%	1.20%	0.06%	
Total	Number	12,867	45,728	8596	1041	63	68,295
	Ratio	18.84%	66.96%	12.59%	1.52%	0.09%	

Table 5
Comparison between VCRO and VCRO previous model.

Model	**Total**	Highest VCRO-values highest conflict severity	**Ratio**
VCRO	68,295	63	0.09%
VCRO previous model	25,421	303	1.2%

ship collision, e.g. the ones presented by Kao et al. (2007) and Goerlandt and Montewka (2015b) incorporate some of these effects. The currently presented model leaves the evaluation of the conflict severity and the occurrence of a near miss but to the expert judgment phase according to Fig. 1. Hence, the VCRO model only suggests cases for further examination, where the hydro-meteorological conditions can be accounted for.

It should also be noted that AIS data contains several deficiencies, with data gaps, missing information and errors rather common, despite the fact that data quality has improved over the years (Felski and Jaskolski, 2013). In such cases, vessel conflicts and related near misses cannot be observed. This presents a possible weakness in using this technique to evaluate maritime traffic safety. However, it is known that maritime accident databases have similar deficiencies, e.g. through underreporting and ambiguity in accident classification (Grabowski et al., 2009; Ladan and Hänninen, 2012). Research into the significance of such uncertainties is left for future work. A related issue is that in detecting near misses, it is advisable to utilize full rate AIS data as certain manoeuvers can be missed when only a subset of the data points is applied (Mestl et al., 2016).

6. Conclusion

In this paper, an enhanced method for detecting near miss ship-ship encounters from AIS data is presented and shown to be effective in reducing the number of encounters requiring further examination by experts. This is useful especially because of the enormous volume of data collected by the AIS. Additionally, the potential for such near miss data to provide further insight into the safety of maritime transportation is discussed.

The Vessel Conflict Ranking Operator (VCRO) recognizes the factors affecting the complexity of ship-ship encounter, and hence ranks collision severity of the encounter. The factors accounted for in the model are the safety distance between the two ships, accounting for the ship domain and thus the vessel size, the relative speed of the ships, the difference between the ship headings, and the available distance for evasive manoeuvring through the MDTC model. The model is constructed based on expert knowledge and models presented in earlier researches.

The discriminant validity tests have been performed and the results show the developed model differentiates effectively in situations representing a different encounter severity level. The tests suggest that the model is adequate for ranking encounters, and for prioritizing the encounters for further expert scrutiny. A comparison with a previously developed VCRO-model shows that the updated model better discriminates encounters of different severity, such than fewer cases will need to be re-contextualized and considered in subsequent expert judgment.

Overall, the contribution of this research is to propose a method detecting possible near miss ship-ship collisions from AIS data without experts involved because some important domain knowledge are taken into consideration in the model such as ship domain, ship intersection angle, etc. Further research should be performed to identify the relationship between near misses and accident involvement, to establish the validity of near misses in assessing maritime traffic safety.

Acknowledgement

The work carried out in this paper has been funded by research funding from PacTrans in University of Washington, and Aalto University. This financial assistance is acknowledged.

Please cite this article as: Zhang, W., et al., An advanced method for detecting possible near miss ship collisions from AIS data. Ocean Eng. (2016), http://dx.doi.org/10.1016/j.oceaneng.2016.07.059

References

Berglund, R., Huttunen, M., 2009. Analysis of Crossing Ship Traffic in the Gulf of Finland VTT-R-10075-08. VTT Technical Research Centre of Finland.

Bukhari, A.C., Tusseyeva, I., Lee, B.-G., Kim, Y.-G., 2013. An Intelligent Real-Time Multi-Vessel Collision Risk Assessment System from VTS View Point Based on Fuzzy Inference System. Expert Syst. Appl. 40, 1220–1230.

Chauvin, C., Lardjane, S., 2008. Decision making and strategies in an interaction situation: collision avoidance at sea. Transp. Res. Part F 11, 259–269.

Chin, H.C., Quek, S.T., 1997. Measurement of traffic conflicts. Saf. Sci. 26 (3), 169–185.

Cockcroft, A.N., Lameijer, J.N.F., 2004. A Guide to Collision Avoidance Rules, 6th ed. Butterworth-Heinemann Ltd., Oxford, UK.

Debnath, A.K., Chin, H.C., 2010. Navigational traffic conflict technique: a proactive approach to quantitative measurement of collision risks in port waters. J. Navig. 63 (1), 137–152.

Debnath, A.K., Chin, H.C., Haque, M.M., 2011. Modelling port water collision risk using traffic conflicts. J. Navig. 64 (4), 645–655.

Faghih-Roohi, S., Xie, M., Ng, K.M., 2014. Accident risk assessment in marine transportation via Markov modelling and Markov chain Monte Carlo simulation. Ocean Eng. 91, 363–370.

Felski, A., Jaskolski, K., 2013. The integrity of information received by means of AIS during anti-collision maneuvering. TransNav – Int. J. Mar. Navig. Saf. Sea Transp. 7, 95–100.

Felski, A., Jaskólski, K., Banys, P., 2015. Comprehensive assessment of automatic identification system (AIS) data application to anti-collision manoeuvring. J. Navig., 1–21. http://dx.doi.org/10.1017/S0373463314000897.

Ferreira, S., Couto, António, 2015. A probabilistic approach towards a crash risk assessment of urban segments. Transp. Res. Part C 50, 97–105.

Fujii, Y., Shiobara, R., 1971. The analysis of traffic accidents – studies in marine traffic accidents. J. Navig. 24, 534–543.

Goerlandt, F., Montewka, J., 2015a. A framework for risk analysis of maritime transportation systems: a case study for oil spill from tankers in a ship-ship collision. Saf. Sci. 76, 42–66.

Goerlandt, F., Montewka, J., 2015b. Maritime transportation risk analysis: review and analysis in light of some foundational issues. Reliab. Eng. Syst. Saf. 138, 115–134.

Goerlandt, F., Montewka, J., Heikki, L., Pentti, K., 2012. Analysis of near Collisions in the Gulf of Finland. In: Berenguer, Grall, Guedes-Soares (Eds.), Advances in Safety, Reliability and Risk Management. Taylor & Francis Group, London, pp. 2880–2886.

Goerlandt, F., Kujala, P., 2014. On the reliability and validity of ship–ship collision risk analysis in light of different perspectives on risk. Saf. Sci. 62 (February), 348–365. http://dx.doi.org/10.1016/j.ssci.2013.09.010.

Goodwin, E.M., 1975. A statistical study of ship domains. J. Navig. 28 (03), 328–344.

Grabowski, et al., 2009. Human and organizational error data challenges in complex, large-scale systems. Saf. Sci. 47 (8), 1185–1194.

Hänninen, M., Kujala, P., 2014. Bayesian network modeling of port state control inspection findings and ship accident involvement. Expert Syst. Appl. 41 (4, Part 2), 1632–1646. http://dx.doi.org/10.1016/j.eswa.2013.08.060.

Hansen, M.G., Jensen, T.K., Lehn-Schiøler, T., Melchild, K., Rasmussen, F.M., Ennemark, F., 2013. Empirical ship domain based on AIS data. J. Navig. 66 (6), 931–940.

Hartigan, J.A., Wong, M.A., 1979. AS 136: A K-Means Clustering Algorithm. J. R. Stat. Soc.: Series C (Applied Statistics) 28 (1), 100–108.

Hauer, E., Gårder, P., 1986. Research into the validity of the traffic conflict technique. Accid. Anal. Prev. 18 (6), 471–481.

Heij, C., Bijwaard, G.E., Knapp, S., 2011. Ship inspection strategies: effects on maritime safety and environmental protection. Transp. Res. Part D 16, 42–48.

HELCOM. 2012. HELCOM Recommendation 33/1 - Unified Interpretation in Relation to Access to and Use of HELCOM AIS. Helsinki Commission, Baltic Marine Environment Protection Commission.

Hydén, C., 1987. The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflict Technique Doctoral. Lund University, Department of Traffic Planning and Engineering, Lund, Sweden.

Kao, S.L., Lee, K.T., Chang, K.Y., Ko, M.D., 2007. A fuzzy logic method for collision avoidance in vessel traffic service. J. Navig. 60 (1), 17–31.

Kujala, P., Hänninen, M., Arola, T., Ylitalo, J., 2009. Analysis of the marine traffic safety in the Gulf of Finland. Reliab. Eng. Syst. Saf. 94 (8), 1349–1357.

Ladan, M., Hänninen, M., 2012. Data Sources for Quantitative Marine Traffic Accident Modelling. Aalto University publication series Science+Technology, Helsinki, Finland, p. 68.

Last, P., Hering-Bertram, M., Linsen, L., 2015. How automatic identification system (AIS) antenna setup affects AIS signal quality. Ocean Eng. 100, 83–89.

Laureshyn, A., Svensson, Å., Hydén, C., 2010. Evaluation of traffic safety, based on micro-level behavioural data: theoretical framework and first implementation. Accid. Anal. Prev. 42 (6), 1637–1646.

Li, S., Meng, Q., Qu, X., 2012. An overview of maritime waterway quantitative risk assessment models. Risk Anal. 32 (3), 496–512.

Mazaheri, A., Montewka, J., Kotilainen, P., Sormunen, O.V.E., Kujala, P., 2015. Assessing grounding frequency using ship traffic and waterway complexity. J. Navig. 68 (1), 89–106.

Mestl, T., Tallakstad, K.T., Castberg, R., 2016. Identifying and analyzing safety critical maneuvers from high resolution AIS data. TransNav – Int. J. Mar. Navig. Saf. Sea Transp. 10 (1), 69–77.

Montewka, J., Hinz, T., Kujala, P., Matusiak, J., 2010. Probability modelling of vessel collisions. Reliab. Eng. Syst. Saf. 95 (5), 573–589.

Montewka, J., Goerlandt, F., Kujala, P., 2012. Determination of collision criteria and causation factors appropriate to a model for estimating the probability of maritime accidents. Ocean Eng. 40 (February), 50–61. http://dx.doi.org/10.1016/j.oceaneng.2011.12.006.

Mou, J.M., Tak, C., van der Ligteringen, H., 2010. Study on collision avoidance in busy waterways by using AIS data. Ocean Eng. 37, 483–490.

Özbaş, B., 2013. Safety risk analysis of maritime transportation: review of the literature. Transp. Res. Rec.: J. Transp. Res. Board 2326, 32–38.

Psaraftis, H.N., Kontovas, C.A., 2014. Ship speed optimization: Concepts, models and combined speed-routing scenarios. Transp. Res. Part C 44, 52–69.

Qu, X., Meng, Q., Li, S., 2011. Ship collision risk assessment for the Singapore Strait. Accid. Anal. Prev. 43 (6), 2030–2036. http://dx.doi.org/10.1016/j.aap.2011.05.022.

Rawson, A., Rogers, E., Foster, D., Phillips, D., 2014. Practical application of domain analysis: port of London case study. J. Navig. 67 (2), 193–209.

Ren, Y.L., Mou, J.M., Yan, Q.X., Zhang, F., 2011. Study on assessing dynamic risk of ship collision, Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation. ASCE, Wuhan, China, pp. 2751–2757. http://dx.doi.org/10.1061/41177(415)346.

Sang, L.Z., Wall, A., Mao, Z., Yan, X.P., Wang, J., 2015. A novel method for restoring the trajectory of the inland waterway ship by using AIS data. Ocean Eng. 110, 183–194.

Silveira, P.A.M., Teixeira, A.P., Soares, C.G., 2013. Use of AIS data to characterise marine traffic patterns and ship collision risk off the Coast of Portugal. J. Navig. 66 (06), 879–898.

Soares, C. Guedes, Teixeira, A.P., 2001. Risk assessment in maritime transportation. Reliab. Eng. Syst. Saf. 74; , pp. 299–309.

Svensson, Å., 1998. A Method for Analysing the Traffic Conflict Technique in a Safety Perspective. Lund University, Department of Traffic Planning and Engineering, Lund, Sweden.

Trochim, W., Donnely, J.P., 2008. The Research Methods Knowledge Base, 3rd ed. Atomic Dog Publishing.

Tsou, M.-C., 2014. Online analysis process on automatic identification system data warehouse for application in vessel traffic service. Proc. Inst. Mech. Eng., Part M: J. Eng. Marit. Environ. . http://dx.doi.org/10.1177/1475090214541426

van Iperen, E., 2015. Classifying ship encounters to monitor traffic safety on the North Sea from AIS data. TransNav – Int. J. Mar. Navig. Saf. Sea Transp. 9 (1), 53–60.

van Iperen, E., 2012. Detection of Hazardous Encounters at the North Sea from AIS Data. In: Proceedings of International Workshop of Next Generation Nautical Traffic Models, 1–12, Shanghai, China.

van Westrenen, F., Ellerbroek, J., 2015. The effect of traffic complexity on the development of near misses on the North Sea. IEEE Trans. Syst., Man, Cybern.: Syst. . http://dx.doi.org/10.1109/TSMC.2015.2503605

Wang, J., Deng, W., Guo, Y.T., 2014a. New Bayesian combination method for short-term traffic flow forecasting. Transp. Res. Part C 43, 79–94.

Wang, N., Meng, X.Y., Xu, Q.Y., Wang, Z.W., 2009. A unified analytical framework for ship domains. J. Navig. 62 (4), 643–655.

Wang, Y., Ma, X.L., Lao, Y.T., Wang, Y.H., 2014b. A fuzzy-based customer clustering approach with hierarchical structure for logistics network optimization. Expert Syst. Appl. 41, 521–534.

Wang, Y., Zhang, J., Chen, X., Chu, X., Yan, X., 2013. A spatial–temporal forensic analysis for inland-water ship collisions using AIS data. Saf. Sci. 57, 187–202.

Wang, Y.Y., Chin, H.C., 2015. An empirically-calibrated ship domain as a safety criterion for navigation in confined waters. J. Navig. . http://dx.doi.org/10.1017/S0373463315000533

Wen, Y.Q., Huang, Y.M., Zhou, C.H., Yang, J.L., Xiao, C.S., Wu, X.C., 2015a. Modelling of marine traffic flow complexity. Ocean Eng. 104, 500–510.

Wen, Y.Q., Huang, Y.M., Zhou, C.H., Yang, J.L., Xiao, C.S., Wu, X.C., 2015b. Modelling of marine traffic flow complexity. Ocean Eng. 104, 500–510.

Williams, M.J., 1981. Validity of the traffic conflict technique. Accid. Anal. Prev. 13, 133–145.

Wu, X., Mehta, A.L., Zaloom, V.A., Craig, B.N., 2016. Analysis of waterway transportation in Southeast Texas waterway based on AIS data. Ocean Eng. 121, 196–209.

Xiao, F.L., Ligteringen, H., van Gulijk, C., Ale, B., 2015. Comparison study on AIS data of ship traffic behavior. Ocean Eng. 95, 84–93.

Zhang, D., Yan, X.P., Yang, Z.L., Wall, A., Wang, J., 2013. Incorporation of formal safety assessment and bayesian network in navigational risk estimation of the Yangtze River. Reliab. Eng. Syst. Saf. 118, 93–105.

Zhang, J.F., Yan, X.P., Chen, C.Q., Sang, L.Z., Zhang, D., 2012. A novel approach for assistance with anti-collision decision making based on the international regulations for preventing collisions at sea. Proc. Inst. Mech. Eng., Part M: J. Eng. Marit. Environ. 226 (3), 250–259.

Zhang, W.B., Goerlandt, F., Montewka, J., Kujala, P., 2015. A Method for detecting possible near miss ship collisions from AIS data. Ocean Eng. 107, 60–69.

Zou, Y., Zhang, Y., Lord, D., 2014. Analyzing different functional forms of the varying weight parameter for finite mixture of negative binomial regression models. Anal. methods Accid. Res. 1, 39–52.